

Article

Killing Educators' Nightmare—Opportunities and Challenges When Using GenAI in Marking an Essay Assignment

Tai Ming Wut ^{*}, Elaine Ah-heung Chan  and Helen Shun-mun Wong

College of Professional and Continuing Education, The Hong Kong Polytechnic University, Hong Kong, China

^{*} Correspondence: edmund.wut@cpce-polyu.edu.hk

Abstract

Generative AI has been used in higher education in teaching and learning. The purpose of this study is to compare the marks generated by GenAI and human educators in essay assignments. The feedback made by GenAI and educators is also examined. A total of 126 submitted assignments from business students in a university were marked by GenAI and by an educator, respectively. It was found that there were small differences in the marks, recorded using a paired sample *t* test, and there was a moderate correlation between educator and GenAI markings. GenAI feedback follows the rubrics closely, and educators are more likely to give a holistic view of the scripts. GenAI marking might be useful in a large class with specific guidelines for an essay format assignment.

Keywords: generative AI; higher education; AI-generated grading; human grading; essay assignment

1. Introduction

The rapid evolution of generative artificial intelligence (GenAI) has brought enormous changes to many industries such as education, business and finance, healthcare and creative industries (Nikolopoulou, 2024). In education, GenAI has brought both benefits and concerns to the sector. It has affected teaching and learning practices, as well as how assessments are designed and conducted (Swiecki et al., 2022). Since 2022, the emergence of advanced language models such as ChatGPT has transformed the landscape of higher education (Lye & Lim, 2024).

Nowadays, academics are overworked and overburdened with different duties such as student enrolments, diversified student needs and research pressure, forcing them to work longer hours and odd hours on weekends (Fauzi et al., 2024), in addition to the marketization of the higher education sector (Taberner, 2018). GenAI can be utilized to replace or complement traditional assessment practices (UNESCO, 2019). Educator assessments in the higher education sector, for example, for essays or group projects, can be labor-intensive and can constitute a substantial workload for academics. Eventually, the instructors may provide inconsistent feedback and subjective grading, which instill unfairness and poor reliability. Upon the development of GenAI, employing AI chatbots for grading students' writing tasks becomes appealing. GenAI may possibly alleviate teachers' workloads by providing objective feedback and comments based on the input rubrics. Therefore, instructors can devote more quality time to providing student services and conducting academic research. However, the inconsistencies between AI-generated assessments and traditional human evaluations remain as concerns (Usher, 2025). The reliability of using GenAI for grading assessments and providing feedback on students'



Academic Editor: Sergio Miranda

Received: 2 January 2026

Revised: 4 February 2026

Accepted: 4 February 2026

Published: 9 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

work is still questionable. This study aims to evaluate and compare the grades and feedback provided by AI chatbots and instructors for students' individual assignments with a word limit of 1000 words for a business subject in a higher education setting. The research questions of this study are as follows:

- (1) How do the grades (marks) generated by GenAI differ from those generated by a human for an essay assignment (i.e., course instructor)?
- (2) How does the quality of feedback given by GenAI differ from the comments provided by a course instructor?

This study is of utmost importance to address the burning issue of instructor overload in higher education, and provides a comprehensive understanding of GenAI's performance in an Asian context where such empirical studies are still scarce. Owing to the inconsistencies and biased personal assessments of instructors, there is a need for utilizing GenAI to provide relief for instructors. Furthermore, the huge workload of instructors is always a burning issue (Defazio et al., 2022). Students' demand for prompt and constructive feedback also overwhelms instructors, which leads to inconsistencies, a lack of quality, and delays in feedback (Edwards et al., 2016). The student learning experience will then be deteriorated. This study is of utmost importance in this AI era.

To further contextualize the relevance of this research, it is important to recognize the global push for digital transformation and automation within higher education. Since universities worldwide are facing a lot of challenges in research, student recruitment, teaching quality, and so on, there is a need to integrate new technologies into their operation so that administrative and academic duties can be streamlined. However, as GenAI is constantly upgraded and becomes more powerful, there have been concerns about academic integrity, plagiarism, and analytical and critical thinking skills (Walter, 2024). Therefore, the aim of this study is to find out the outcomes and performance of using GenAI for grading assessments so that more implications can be drawn for higher education to incorporate GenAI into teaching and learning practices. At the same time, educators can also gain a better understanding about how to leverage GenAI to enhance students' learning and utilize it for grading student assessments as well. Finally, the results can also serve as a reference point for tertiary educators to develop appropriate strategies using combinations of both technological advancement and human involvement.

2. Literature Review

2.1. Generative Artificial Intelligence (GenAI) in Higher Education

With the development of GenAI, AI chatbots are commonly adopted because of their automated conversational agents powered by large language models (LLMs) (Okonkwo & Ade-Ibijola, 2021). These chatbots can process natural language and engage in dialog so as to perform some assessment-related tasks, such as grading and giving feedback and comments (Okonkwo & Ade-Ibijola, 2021). AI chatbots can act as a virtual teaching assistant, affecting how students learn, how instructors teach, and how assessment is designed and conducted (Swiecki et al., 2022). For students, these tools not only provide support in coursework, subject revision, and concept clarification, but also provide immediate, personalized feedback for immediate improvement (Usher, 2025; Usher & Amzalag, 2025). This enables students to achieve a better understanding of the subjects and a more capable performance in assignments and assessments. Nevertheless, the outstanding outcomes generated by GenAI have brought up several issues regarding students' academic integrity and plagiarism, which are always the most critical issues in tertiary education (Moorhouse et al., 2023; Zhao et al., 2024). This further leads to the importance of educators in the university context (Klyshbekova & Abbott, 2024). Eventually, the emergence of GenAI has affected assessment design such that that students cannot easily make use of GenAI to

complete their assessment with ease. For example, an oral defense may be needed after the students have submitted their essay (Mulder et al., 2023).

On the contrary, GenAI provides benefits to educators in terms of the preparation of teaching materials and designing in-class assessments and tests, which may reduce the daily workload of the instructors. Other than that, GenAI can also facilitate the more efficient and consistent grading of assessments because it does not feel fatigue or get tired like humans. Therefore, GenAI does contribute to a more efficient and consistent assessment in grading, further transforming the educational landscape. AI chatbots are recognized for their potential to respond to individual learners' needs, providing personalized support (Labadze et al., 2023). Building on this capacity, generative AI (GenAI) now presents innovative approaches for evaluating student work, with emerging research exploring its application in grading and feedback (Zhao et al., 2024; Usher, 2025). GenAI can respond to each individual learner's needs to provide innovative approaches to grading students' work (Labadze et al., 2023). GenAI can also assist in a wide range of academic activities, such as responding to students' enquiries on assignments, suggesting useful and relevant resources, and answering all enquiries from students instantly (Usher & Amzalag, 2025). Ultimately, these tools help develop a learning context which is easily accessible to students, being more responsive to individual students' needs (Walter, 2024). However, the use of GenAI by students in their submitted assignments might demand educators to judge whether the work is done by a human or by GenAI. To conclude, the adoption of GenAI in the education sector does bring both opportunities and challenges to educators.

2.2. GenAI and Assessment

The evolution of GenAI has affected educational assessments in the tertiary education sector in terms of their design and execution. The high-quality responses from GenAI lead to its popularity for grading academic essays from students (Zhao et al., 2024). As such, students can easily manipulate GenAI to help complete their assignments within a short period of time, for example, in few minutes. On the contrary, concerns about academic integrity and dishonesty have been raised. Notably, the world's 50 top-ranking tertiary education institutions have established a set of guidelines and policies to deal with issues such as plagiarism, acknowledgement of using GenAI, detection of Gen AI usage, and assessment design (Moorhouse et al., 2023).

Consequently, research (e.g., C. K. Y. Chan & Colloton, 2024; Furze, 2025; Khlaif et al., 2025) has discussed the redesign of assessments to ensure that learning outcomes can be achieved even under the influence of GenAI. Without a doubt, GenAI has rapidly transformed educational assessment practices in the tertiary education sector. It can be applied in automated grading and intelligent tutoring systems, as well as for giving tailor-made feedback, self-directed learning support, and scalable evaluation mechanisms (Adiguzel et al., 2023). GenAI can also facilitate a series of assessment activities such as automated essay grading, formative feedback generation, the detection of plagiarism and adaptive testing (Lye & Lim, 2024). Nevertheless, GenAI is still in its infancy in terms of delivering contextually relevant feedback, especially for tasks assessing students' analytical reasoning, critical thinking or decision-making skills (Usher, 2025). GenAI's capacity to diagnose discipline-specific content remains limited, and its effectiveness also fluctuates a lot with different task types (Usher, 2025). These findings suggest the need for the careful and critical adoption of GenAI in current assessment practices in the tertiary education section. Human monitoring is highly needed to maintain academic standards and fairness. Without a doubt, GenAI does help provide instant feedback to students for revision and writing improvement; it also contributes to the instructors' side of grading assessments, while creating new challenges in assessment design.

2.3. GenAI and Grading

Apart from assessment design, GenAI can also assist in grading students' work. The current literature focuses on how GenAI can help students to learn, while only limited studies discuss the benefits for educators in teaching and learning, especially grading assessments. There are many advantages of using GenAI for grading assessments. By giving pre-designed assessment criteria and rubrics, GenAI can evaluate and assess students' work based on this information. Moreover, grades with detailed feedback and suggestions can also be provided at the end, which enhances instructors' grading workload. A key argument for using GenAI in grading is its potential to provide timely, immediate, and consistent feedback, allowing students to understand how to improve their work (Labadze et al., 2023). To help educators, GenAI can help in automating the grading of different types of student work, from straightforward multiple-choice tests to complicated writing essays (Nikolopoulou, 2024). Therefore, GenAI can save instructors time and effort in grading because it can function effectively and efficiently even under a huge number of assignments. GenAI can automate grading tasks so that instructors can devote more time to doing research, curriculum development, and meaningful student engagement. Moreover, GenAI can generate instant and personalized feedback, enhancing students' learning by providing timely insights and opportunities for improvement. Zhao et al. (2024) emphasized that GenAI supports instructors by reducing workload and ensuring grading consistency. Several studies, including those by Lu et al. (2024), Okonkwo and Ade-Ibijola (2021), and Usher (2025), have compared AI-generated and human grading. Their findings indicate that GenAI delivers a reliable consistency and comparable accuracy, particularly in low-stakes or high-volume assessments. Unlike humans, AI can perform its tasks with no bias and a good accuracy and consistency because it does not feel fatigue or get bored (Fischer, 2025). However, GenAI does not perform well when the grading tasks require higher-level thinking skills such as creativity, analytical reasoning and critical thinking. For example, Usher (2025) has conducted a study to compare the three grading outputs from AI, peers, and instructors for group projects. It was found that AI tended to grade higher than human instructors. Perhaps humans do their marking with their perceptions towards students, which are not totally unbiased and fair. In addition, AI's feedback is generic and sometimes does not align with specific assessment criteria and the subject context. GenAI also performed poorly when grading complicated and qualitative student work (Okonkwo & Ade-Ibijola, 2021; Lu et al., 2024; Usher, 2025). In a study conducted by Lu et al. (2024), it was found that there was a moderate to strong consistency between AI-generated grades and human-generated grades in students' writing works. Even though GenAI can be easily adopted into assessment practices, it also brings a lot of issues regarding its accuracy, reliability, consistency, and pedagogical validity (Lu et al., 2024). Occasionally, GenAI may misunderstand or even misinterpret students' work, leading to incorrect or unhelpful feedback given (H. C. B. Chan & Hu, 2023). Therefore, merely using GenAI for marking may not be an ideal option. Consequently, a sole reliance on AI for summative or formative assessment is not advisable. Instead, a hybrid model that strategically combines the efficiency of AI with the professional judgment of educators is essential for upholding assessment integrity, ensuring fairness, and supporting meaningful learning outcomes.

2.4. Affordances and Challenges of GenAI in Education

Since 2023, the affordances of GenAI tools adopted in education and assessment have been widely researched. Several affordances were identified, including (1) the capability of GenAI to give timely help, and its accessibility for remote learning; (2) its ability to provide personalized learning recommendations; (3) the use of automation to provide support and responses to users' prompts; and (4) its ability to serve as a facilitator in the

enhancement of the conversational abilities of users (Wang et al., 2024). While many benefits of using GenAI in higher education were documented in the early stage, the challenges and issues of adopting GenAI in education were discussed. The ethical implications of using GenAI in higher education were raised by the authors of the UNESCO guidance, Miao and Holmes (2023). Usher (2025) also highlighted that GenAI lacks explainability and transparency, and that no one knows how it operates and processes its outputs. Wang et al. (2024) also asserted five categories of challenges: the risks of academic integrity when students use AI for assignments without proper disclosure, response errors and bias, an over-reliance on GenAI, digital divide issues, and privacy and security issues. An overreliance on GenAI may affect students' "real" learning because they blindly adopt answers without understanding the topic (Nah et al., 2023; Wang et al., 2024). Such developments may further inhibit the advancement of critical thinking, problem-solving skills, and independent learning capacities among students in higher education (Usher, 2025). The adoption of GenAI for student assessment also raises significant concerns regarding the privacy and security of student data. It often requires access to sensitive personal and academic information to function effectively; there is an increased risk of data breaches, unauthorized access, and the potential misuse of information (Lye & Lim, 2024).

2.5. Research Gap

Although more research has been conducted examining how GenAI assists human grading, the findings are far from consistent. More empirical studies examining the impact of AI on assessment gradings should be conducted to enhance our understanding of the potential of manipulating AI for grading so that consistent and unbiased grading can be possible. Moreover, AI should not be viewed as a replacement for humans in education (Usher, 2025). Instead, it should be considered as a digital teaching assistant for the instructor, so as to reduce the instructor's workload. Therefore, a collaboration strategy combining humans and AI would be feasible and practical for the education industry (Usher, 2025). Finally, it is important to note that the majority of existing research in this domain has been predominantly conducted within the context of the United States, with comparatively limited empirical investigations undertaken in Asian countries. This geographic concentration of studies raises concerns regarding the external validity and generalizability of the findings, as cultural, educational, and institutional differences across regions may significantly influence the applicability of the results. Expanding research efforts to include diverse Asian contexts is therefore essential to ensure a more comprehensive and globally representative understanding of the phenomena under investigation.

This study addresses these gaps by empirically comparing GenAI and human grading within a business education context in Hong Kong. It investigates both quantitative marks and qualitative feedback to contribute evidence from an under-researched region and inform the development of practical collaboration models.

Thus, the research hypotheses are listed as follows:

Hypothesis 1. *The mark for an essay assignment marked by generative AI is different compared to the assignment marked by a human being.*

Hypothesis 2. *Comments given on an essay assignment marked by a generative AI are different compared to the assignment marked by a human being.*

3. Materials and Methods

Business students attended a compulsory course in their third year of studies. The course's name is "Environmental, Social and Governance Management". The purpose of the course is to provide a foundation for students in environmental sustainability, social

responsibility and governance problems. The contact hours are 39 h in total and consist of two hours of lectures and a one-hour tutorial per week. There are in-class exercises, individual assignments, a test, and finally a group project. There is no examination at the end of the semester. The medium of instruction for the course is English, and most students are a second language learner.

Regarding an individual assignment, students were asked to write a one-thousand-word essay on one of the seventeen sustainable development goals assigned to them. A brief guide was sent to the students in week three. Students were expected to submit their work in week five. Students are expected to report on the targets and current achievements of the assigned sustainable development goal. The challenges of reaching the sustainable targets and their recommendations are needed. Students need to provide all references and in-text citations.

There were 126 submitted assignments. With an effect size of 0.5, the sample size requirement for the paired sample *t* test is 45, and the level of significance equals 0.05, determined using the G*Power 3.1.9.7 (Faul et al., 2007). The same assignment was marked by GenAI and by the subject lecturer, respectively. Ethics approval was obtained from the research committee of the institution (RC/ETH/H446). Written informed consent was obtained from the students before the study. Before the marking, the whole question, the requirements of the individual assignment, the assessment rubrics of the assignment and the grade-to-mark conversion were given as input to the GenAI. Co-pilot with ChatGPT 5 was used in this study. According to the answers provided by Co-pilot, Co-pilot is provided by Microsoft's AI companion design to help users to complete tasks. It is a knowledge partner as well as an assistant. Co-pilot works as an AI chatbot, interacting with people who use it.

4. Results

The mean mark given by the human being was 70.37, with a standard deviation of 11.673, whereas that given by GenAI was 67.21, with a standard deviation of 10.050. It is not surprising that the spread of the mark was greater for human marking. The correlation between human and GenAI marking was 0.420, which is moderate in strength. Using a paired sample *t* test to test the hypothesis, the research hypothesis is supported, with a *p*-value smaller than 0.01. The effect size of Cohen's *d* is 0.268. This means that there is a difference between GenAI and human marking, with small differences.

Regarding the comments for the second hypothesis, the GenAI marking was more comprehensive compared to human marking. Human beings tend to present the most important comment: for example, "please pay attention to the in-text citations, some are missing." The length of human beings' comments was usually shorter. Usually, one or two phrases were written by human markers. In contrast, GenAI provided detailed comments according to the rubrics given. The assessment rubrics were analysis, reasoning and creativity (25%); evidence of information search and supporting evidence of arguments (25%); theory and concept application (25%); and presentation (25%). The strengths of the students are listed: for example, "you have logical reasoning," "you have provided a well-structured and comprehensive essay, covering all required sections of the assignment," "when you are applying theories or concepts, name them and show how they inform your arguments", and "You need to provide linking sentences to link up paragraphs. Some paragraphs are too long" for a presentation. In contrast, a human being marker tends to give a holistic statement. For example, "pay attention to the referencing format and in-text citations."

5. Discussion and Conclusions

The correlation between GenAI marking and human marking is almost the same as the result from Usher (2025). However, there is only a small difference between GenAI and human marking in individual assignments. However, there is a larger difference between GenAI and humans when marking group projects than in the previous study (Usher, 2025). This might be due to the selection of topics being free-handed, and students would like to express their best to the educators. In contrast, there are some specific guidelines in the individual assignments, such that GenAI can follow those guidelines. The results from Usher (2025) show that higher marks were given by GenAI marking, which is in contrast to our results.

GenAI's marking is closely related to human marking, and there is only a small difference between GenAI and human marking in the essay assignment. It certainly alleviates teachers' working burden. It is anticipated that both markings would be closer with specific guidelines for the assignment.

It seems that it is possible to use GenAI in marking technically, but educators need to pay attention to the comments GenAI makes in the assignment. There could be a difference in the grades given by GenAI and a human marker. It would be needed to cross-check some samples to maintain consistency. There are limitations in the study. Business assignments in essay format were used in our research. Some other disciplines and formats can be used in further research.

The adoption of GenAI is likely to shape the future landscape of higher education. However, GenAI tools can exceed humans by processing huge amounts of data efficiently; their accuracy, fairness, and effectiveness in assessing students' critical thinking and writing skills remain a challenge. Moreover, instead of using AI to replace human judgment, a human–AI collaboration model in assessment may be feasible (Chang et al., 2023; Usher, 2025). The advantages and disadvantages of different grading approaches may be identified for better learning efficiency. By doing so, the study can provide educators, researchers, and policymakers with a comprehensive understanding of the challenges and opportunities when utilizing AI for assessing student work, and can propose a human–AI collaboration model for this field. Finally, directions for future research and practices in AI-generated assessments could be provided. There are limitations for the study. The sample size is relatively small, and one human marker was involved in the marking process. The result may not be generalizable. The limitations of using GenAI for student assessments can be better explored and recognized using theoretical models. AI developers may also further improve their large language models to enhance the effectiveness of AI-assessed grading and feedback.

Author Contributions: Conceptualization, T.M.W.; methodology, T.M.W., software, T.M.W., validation, T.M.W., formal analysis, T.M.W., investigation, T.M.W., resources, T.M.W., data curation, T.M.W., writing-original draft preparation, T.M.W., writing-review and editing, E.A.-h.C., visualization, E.A.-h.C., supervision, H.S.-m.W., project administration, H.S.-m.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethics approval was obtained from the Research Committee of The Hong Kong Polytechnic University College of Professional and Continuing Education, approval code is RC/ETH/H446 and approval date is 1 December 2025.

Informed Consent Statement: Written informed consent was obtained from the students before the study.

Data Availability Statement: Data is available from corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Adiguzel, T., Kaya, M. H., & Cansu, F. K. (2023). Revolutionizing education with AI: Exploring the transformative potential of ChatGPT. *Contemporary Educational Technology*, 15(3), ep429. [CrossRef]
- Chan, C. K. Y., & Colloton, T. (2024). Redesigning assessment in the AI era. In *Generative AI in higher education*. Chapter 4. Routledge. [CrossRef]
- Chan, H. C. B., & Hu, B. (2023, November 27–December 1). *Grading generative AI-based assignments using a 3R framework*. 2023 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE) (pp. 1–5), Auckland, New Zealand. [CrossRef]
- Chang, D. H., Lin, M. P. C., Hajian, S., & Wang, Q. Q. (2023). Educational design principles of using AI chatbot that supports self-regulated learning in education: Goal setting, feedback, and personalization. *Sustainability*, 15(17), 12921. [CrossRef]
- Defazio, D., Kolymiris, C., Perkmann, M., & Salter, A. (2022). Busy academics share less: The impact of professional and family roles on academic withholding behaviour. *Studies in Higher Education*, 47(4), 731–750. [CrossRef]
- Edwards, D. B., Jr., & Loucel, C. (2016). The EDUCO program, impact evaluations, and the political economy of global education reform. *Education Policy Analysis Archives*, 24(92). [CrossRef]
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39, 175–191. [CrossRef]
- Fauzi, M. A., Rahamaddulla, S. R., Lee, C. K., Ali, Z., & Alias, U. N. (2024). Work pressure in higher education: A state of the art bibliometric analysis on academic work–life balance. *International Journal of Workplace Health Management*, 17(2), 175–195. [CrossRef]
- Fischer, L. (2025). The promise and pitfalls of AI in academic grading. *Technology in Education Review*, 12(3), 77–89.
- Furze, L. (2025, August 18). *Five principles for rethinking assessment with GenAI*. Available online: <https://leonfurze.com/2025/08/18/five-principles-for-rethinking-assessment-with-gen-ai/> (accessed on 1 December 2025).
- Khlaif, Z. N., Alkhouk, W. A., Salama, N., & Abu Eideh, B. (2025). Redesigning assessments for AI-enhanced learning: A framework for educators in the generative AI era. *Education Sciences*, 15(2), 174. [CrossRef]
- Klyshbekova, M., & Abbott, P. (2024). ChatGPT and assessment in higher education: A magic wand or a disruptor. *Electronic Journal of e-Learning*, 22(2), 30–45. [CrossRef]
- Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: Systematic literature review. *International Journal of Educational Technology in Higher Education*, 20(1), 56. [CrossRef]
- Lu, Q., Yao, Y., Xiao, L., Yuan, M., Wang, J., & Zhu, X. (2024). Can ChatGPT effectively complement teacher assessment of undergraduate students' academic writing? *Assessment & Evaluation in Higher Education*, 49(5), 616–633. [CrossRef]
- Lye, C. Y., & Lim, L. (2024). Generative artificial intelligence in tertiary education: Assessment redesign principles and considerations. *Education Sciences*, 14(6), 569. [CrossRef]
- Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO.
- Moorhouse, B. L., Yeo, M. A., & Wan, Y. (2023). Generative AI tools and assessment: Guidelines of the world's top-ranking universities. *Computers and Education Open*, 5, 100151. [CrossRef]
- Mulder, R., Baik, C., & Ryan, T. (2023). *Rethinking assessment in response to AI*. Melbourne Centre for the study of Higher Education. The University of Melbourne.
- Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, 25(3), 277–304. [CrossRef]
- Nikolopoulou, K. (2024). Generative artificial intelligence in higher education: Exploring ways of harnessing pedagogical practices with the assistance of ChatGPT. *International Journal of Changes in Education*, 1(2), 103111. [CrossRef]
- Okonkwo, C. W., & Ade-Ibijola, A. O. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, 2, 100033. [CrossRef]
- Swiecki, Z., Khosravi, H., Chen, G., Martinez-Maldonado, R., Lodge, J. M., Milligan, S., Selwyn, N., & Gašević, D. (2022). Assessment in the age of artificial intelligence. *Computers and Education: Artificial Intelligence*, 3, 100075. [CrossRef]
- Taberner, A. M. (2018). The marketisation of the English higher education sector and its impact on academic staff and the nature of their work. *International Journal of Organizational Analysis*, 26(1), 129–152. [CrossRef]
- UNESCO. (2019, May 16–18). *Planning education in the AI era: Lead the leap*. International Conference on Artificial Intelligence and Education, Beijing, China.
- Usher, M. (2025). Generative AI vs. instructor vs. peer assessments: A comparison of grading and feedback in higher education. *Assessment & Evaluation in Higher Education*, 50(6), 912–927. [CrossRef]
- Usher, M., & Amzalag, M. (2025). From prompt to polished: Exploring student–chatbot interactions for academic writing assistance. *Education Sciences*, 15(3), 329. [CrossRef]

- Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education*, 21(1), 15–29. [\[CrossRef\]](#)
- Wang, N., Wang, X., & Su, Y. S. (2024). Critical analysis of the technological affordances, challenges and future directions of Generative AI in education: A systematic review. *Asia Pacific Journal of Education*, 44(1), 139–155. [\[CrossRef\]](#)
- Zhao, J., Chapman, E., & Sabet, P. G. P. (2024). Generative AI and educational assessments: A systematic review. *Educational Research and Practice*, 51, 124–155. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.